

# **Spatial Prediction of Landslide Hazard Using Discriminant Analysis and GIS**

GIS in the Rockies 2000 Conference

*Peter V. Gorsevski*

*Paul Gessler*

*Randy B. Foltz*

## **Abstract**

Environmental attributes relevant for spatial prediction of landslides triggered by rain and snowmelt events were derived from digital elevation model (DEM). Those data in conjunction with statistics and geographic information system (GIS) provided a detailed basis for spatial prediction of landslide hazard. The spatial prediction of landslide hazard in this paper is based on discriminant analysis. Discriminant analysis is a multivariate technique that can be used to build rules that can classify elements or observations successfully between stable and unstable areas. The discriminant rule would show how to take into account the relative risks of making errors of misclassification. Those general rules allow managers to consider that errors in one direction may be much more costly than errors in the other direction.

## **Keywords**

Landslides, Landslide Hazard, Slope Stability, Geographic Information System, Spatial Prediction, Multivariate Models, Discriminant Analysis

---

## **Introduction**

Human activities, such as deforestation and urban expansion, accelerate the process of landslides (Chung et al., 1995). These activities can affect the process of erosion and sedimentation, with harmful impacts on aquatic habitat. The dominant erosion and sedimentation processes are due to surface erosion or mass failure. Mass failure is a gravity-driven process that occurs when the shear strength of a soil mass is overcome by the shear stresses acting against it. Mass failure is

triggered by rain and snowmelt events and it is the most important cause of sedimentation in areas of steep slopes and unstable soils. Concern about mass failure calls for improved forest management practices in all forestlands where human activities have taken place.

In recent years the use of GIS for landslide hazard modeling has increased because of the development of commercial systems, such as Arc/Info (ESRI) and the quick access to data obtained through Global Positioning System (GPS) and remote sensing. The use of multivariate statistics with GIS has been studied for a long time (Carrera et al., 1983, 1991, 1992, 1995). At the beginning stage of landslide hazard modeling large grid cells with a ground resolution of 200 by 200 m were used. Although the method based on spatial correlation has not undergone major changes, the basic modeling element (cell size) and the tools for modeling have improved significantly.

The statistical model developed by Carrera et al., is built up in a "training area" where the spatial distribution of the landslides is well known. After the model is extended to the entire study area ("target area"), it is assumed that factors that cause slope failure in the study area are the same as these in the "training area". The landslide hazard modeling is achieved by discriminant analysis and multiple regression.

More recently, Mark and Ellen (1995) applied logistic regression for predicting sites of rainfall-induced shallow landslides that initiate debris flow. In their study, statistics were used to determine the best correlation between mapped debris flow sources and physical attributes thought to influence shallow landsliding.

Gorsevski et al. (2000), also applied logistic regression for spatial prediction of landslide hazard. Logistic regression was used to predict the probability of occurrence in tandem with the receiver operator characteristic (ROC) curve as a measure of performance of a predictive rule. The ROC curve is a plot of the probability of having true positive identified landslides versus the probability of false positive identified landslides as the cut-off probability varies. ROC curve was used when decision making was done with uncertainty. The performance of ROC curve may be systematically varied by choice of a decision parameter that adjusts the trade-off between the proportion of pixels correctly identified and those incorrectly identified.

However, few of these studies integrate continuous and discrete variables for statistical analysis, and multivariate methods are not often used for exploratory data analysis. Integrating continuous and discrete variables is done by converting the discrete variables to dummy variables and joining the dummy variables with the continuous variables. When multivariate methods are used for exploratory data analysis the advantages are that fewer new variables can be created, it is easier to locate or identify abnormalities, and it is easier to check assumptions that may be required for certain statistical analysis to be valid. For this study the development of the quantitative hazard model uses continuous and discrete variables and Principle Component Analysis (PCA) as a first step for screening the data.

In this paper we focus on Rocky Point, a small watershed of the Clearwater River Basin in central Idaho, where a total of 23 landslides were recorded (Figure 1). The Rocky Point

watershed is located in central Idaho in the Clearwater River Basin and occupies approximately 4230 hectares (Figure 1). The highly dissected mountainous topography of the Rocky Point watershed is typical for Idaho's Clearwater River Basin. Elevation in the Rocky Point watershed ranges from 1097 m to 1889 m and slopes vary between 0 and 44 degrees. Precipitation averages about 1100 mm annually.

The trigger for mass failure is blamed on the landslide events that occurred during the winter of 1995-96 following heavy rains, snowmelt, and high river flow in the Clearwater National Forest. The landslide data were collected through a combination of photo interpretation and field inventory (McClelland et al., 1997). DEM with TAPES-G software (Gallant et al., 1996) was used to generate quantitative environmental attributes. Environmental attributes included: elevation, slope, aspect, profile curvature, plan curvature, tangent curvature, flow path and upslope area (Table 1).

Attribute	Definition	Significance
Elevation	Height above sea level	Climate, vegetation, potential energy
Slope	Change in elevation divided by horizontal distance	Overland and subsurface flow velocity and runoff rate, precipitation, vegetation, geomorphology, soil water content, land capability class
Aspect	Slope azimuth	Solar insolation, evapotranspiration, flora and fauna distribution and abundance
Profile curvature	Slope profile curvature	Flow acceleration, erosion/deposition rate, geomorphology
Plan Curvature	Contour curvature	Converging, diverging flow, soil water content, soil characteristics
Tangent Curvature	Curvature of line formed by intersection of surface with plane normal to flow line	Erosion/deposition
Flow Path	Distance from watershed divide to the point of interest	Erosion rates, sediment yield, time of concentration
Upslope slope	Mean slope of upslope area	Runoff velocity, potential energy

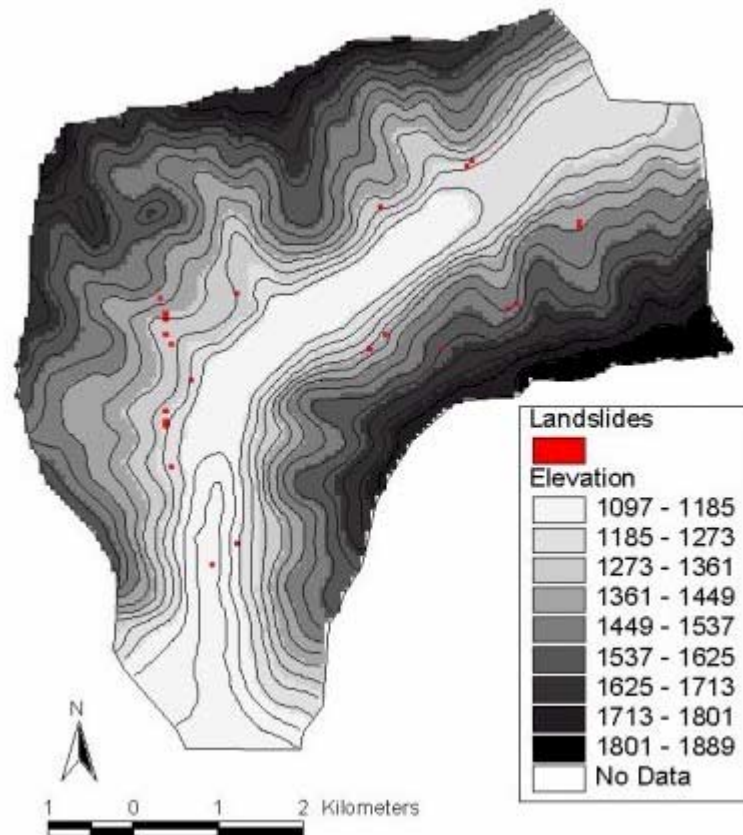
**Table 1.** Environmental attributes derived from digital elevation model.

In the last several years, tools and techniques for developing quantitative hazard models and prediction of landslide hazard have advanced considerably (Carrera et al., 1991; Montgomery and Dietrich, 1994; Mark and Ellen, 1995). Multivariate methods have been applied to various disciplines such as soil-landscape modeling, and more recently in environmental science (Carrera, 1983; Gessler et al, 1995; Gorsevski et al, 2000). In this study the spatial prediction of landslide hazard is based on discriminant analysis. Discriminant analysis is a multivariate technique that is used to build rules that can classify landslide hazard into appropriate class.

Discriminant analysis is similar to regression analysis except that the dependent variable is categorical rather than continuous. We implement discriminant rules that take into account the relative risks of making errors of misclassification.

## Methods

All environmental attributes in this study were continuous, except the aspect and the parent material. For those two environmental attributes dummy variables for each of the categories of the nominal variables were created. A set of 26 attributes including the dummy variables was acquired for entire study area. The mapping unit used to describe the difference between adjacent units across the study area was set to a cell size of 60 by 60 m. A total of 5 % of the mapping units was sampled for constructing the subset (training area) to be used for the discriminant analysis.



**Figure 1.** Distribution of landslides over the Rocky Point watershed

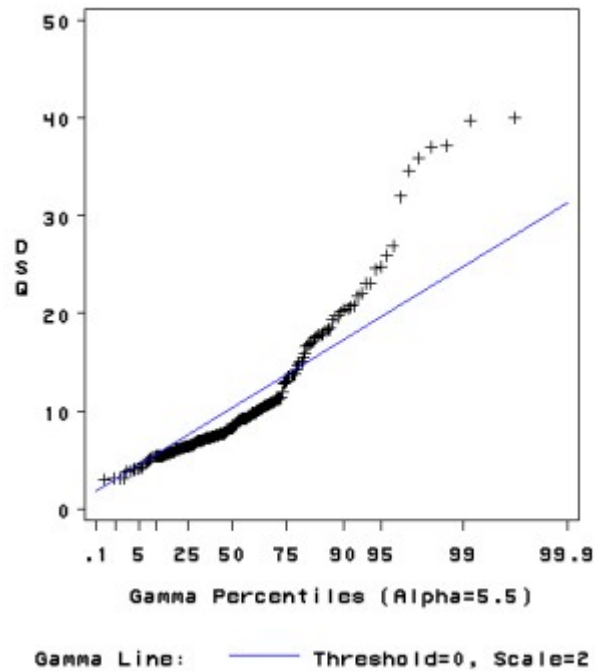
As a first step in this study, a principal component analysis (PCA) was performed on the subset data to help analyze the multivariate data set. The set of 26 attributes was transformed into a new

set of 11 attributes called principle component scores. The number of principal components for the new data set was selected through a combination of scree plot and at least 80% of total variability in the original variables. A multivariate normal plot was used to check whether the newly created principle component scores were distributed according to a multivariate normal distribution. When the predictor variables have a multivariate normal distribution, then the discriminant rules are known to be the best at discriminating (Johnson, 1998).

Nearest neighbor discriminant analysis was used to classify presence and absence of landslides using the principle component scores. The estimate of probabilities of misclassification was based on cross-validation. Cross-validation method removes each observation vector from the calibration data set at a time, forms the discriminant rule based on all the remaining data to classify the removed observation, and notes whether the observation is correctly classified. When developing the discriminant rules for both populations (presence and absence of landslides) we take into account the relative risks of making errors of misclassification. Errors in one direction might be more serious than errors in the other direction. By assigning relative costs of equal or unequal misclassification of these two kinds of errors the development of discriminant rules for both populations show various outcomes of the final classification. When we assign relative costs of equal misclassification than the errors in both directions are equal. Conversely, when we assign relative costs of unequal misclassification the errors in one direction is larger. It is apparent that the probability of making an error in one direction can be reduced, but it is also apparent that reduction of probability of making an error in one direction will cause an increase in the probability of making an error in the other direction.

## **Discussion**

The multivariate normal plot shown in Figure 2 was constructed from the principles component scores, and it is not distributed normally. If the normality assumption is not satisfied, the nearest neighbor analysis is used. Nearest neighbor discriminant analysis is a nonparametric discrimination procedure, which depends on the Mahalanobis distance between pairs of observation vectors (Johnson, 1998). The new observation is classified by finding the closest observation (Mahalanobis distance) in the calibration data set, and assigning the new observation to the group from which the majority of observation's nearest neighbor came. For this analysis the nearest neighbor ( $k$ ) was set to 5, for classifying the new observations. Although we used a few different options for the nearest neighbor ( $k$ ), for our dataset ( $k$ ) was set to 5 because it provides the best discriminant rule.



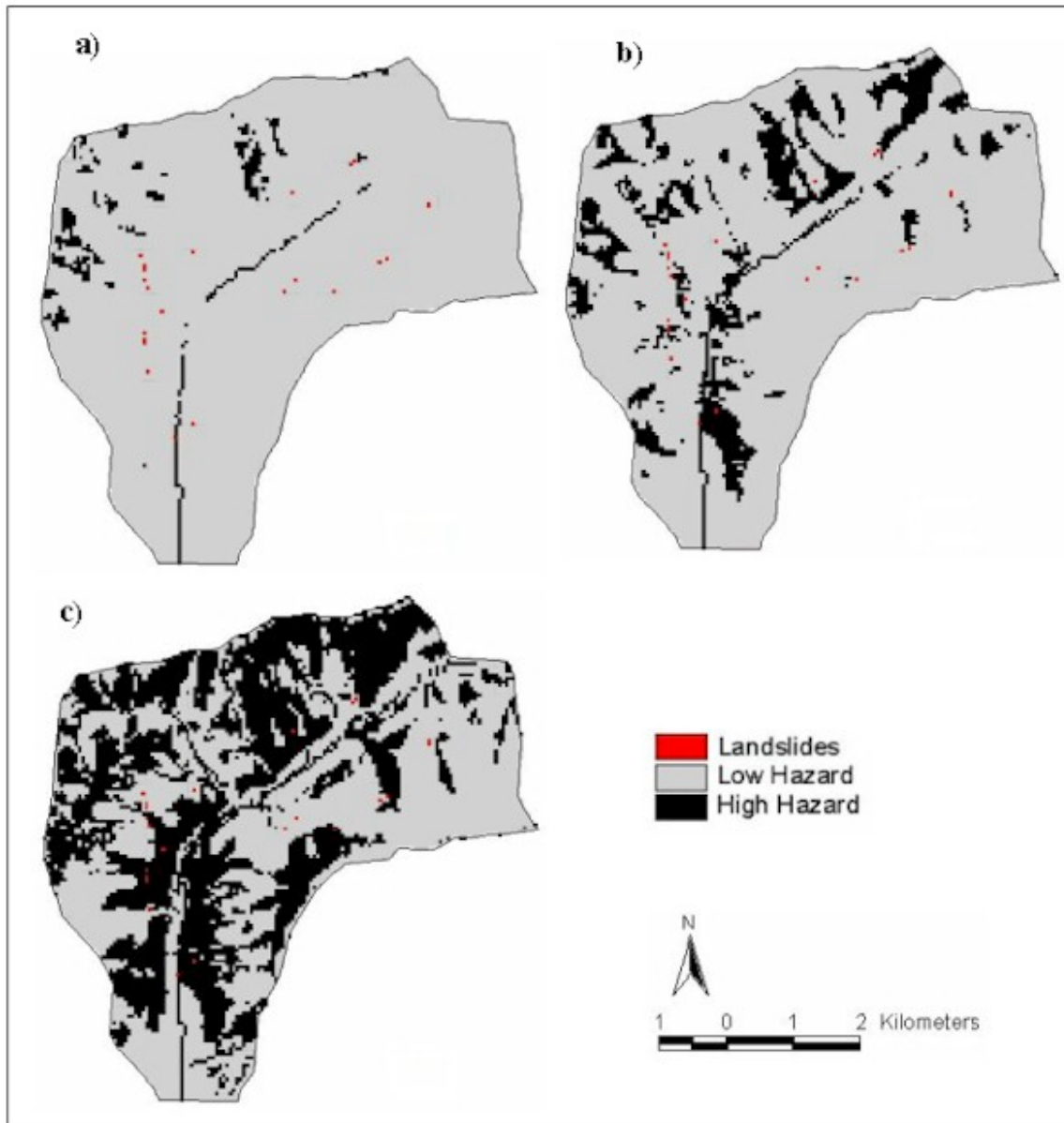
**Figure 2.** Multivariate normal probability plot for the principles component scores

The classification results in the calibration data were classified by the cross-validation method. The summary matrices for the cross-validated estimates using equal and unequal costs are shown in Table 2. The summary in the table shows three different cases of the calibration data. The three different cases show how each landslide and nonlandslide data would be classified by the discriminant rule. The difference between the discrimination rule is associated with the cost function. For example, Case I shows a situation when it is nine times as costly to classify a nonlandslide cell as a landslide, than it is to classify a landslide cell as nonlandslide. Case II shows a situation when it is four times as costly to classify a nonlandslide cell as a landslide, and lastly Case III shows a situation when equal cost is used for both populations. Although the cost function is different, it is assumed that prior probability for developing the discriminant rule are equal.

		Absent	Present
Case I	Absent	98.20%	1.80%
	Present	90.91%	9.09%
Case II	Absent	86.83%	13.17%
	Present	68.18%	31.82%
Case III	Absent	68.26%	31.74%
	Present	36.36%	63.64%

**Table 2.** Percent Classified into presence or absence of landslides

From the results Case I shows 98.2% of the observations are correctly classified as nonlandslide data, and 9.09% of the observations are correctly classified as landslide data. Case II shows 86.83% of the observations are correctly classified as nonlandslide data, and 31.82% of the observations are correctly classified as landslide data. Finally, Case III shows 68.26% of the observations are correctly classified as nonlandslide data, and 63.64% of the observations are correctly classified as landslide data. The prediction of landslide hazard is shown in the three maps derived by the discriminant rules mentioned above, and Figures 3 (a) (b) and (c) illustrates all three cases of landslide hazard.



**Figure 3.** Hazard maps generated with different cost functions

## Conclusion

This paper has outlined the process of mapping landslide hazards using GIS, discriminant analysis, taking into account the relative risks of making errors of misclassification. Decision makers can choose the level of risk they are prepared to accept based on the final outcome of several different alternatives using different cost functions. Reduction of the probability of



making an error in one direction is possible when there are only two populations (landslide and nonlandslide). However, reducing the probability of making an error in one direction causes an increase in the probability of making an error in the other direction. These discriminant rules for minimizing the error in one direction allow managers to use prior information (knowledge) based on experience and heuristic knowledge.

## **Acknowledgements**

The research and work presented in this paper was supported through data provided by the Clearwater National Forest and the Rocky Mountain Research Station in Moscow, Idaho.

## **References used**

- Carrera A., 1983. Multivariate Models for Landslide Hazard Evaluation, *Mathematical Geology*, Vol. 15, No. 3, 1983
- Carrera A, M. Cardinali, and F. Guzzetti 1992. Uncertainty in Assessing Landslide Hazard and Risk. *ITC Journal*, No. 2, pp. 172-183.
- Carrera A., M. Cardinali, R. Detti, F. Guzzetti, V. Pasqui, and P. Reichenbach 1991. GIS Techniques and Statistical Models in Evaluating Landslide Hazard. *Earth Surface Processes and Landforms*, Vol. 16, 427-445 (1991)
- Carrera A., M. Cardinali, F. Guzzetti, and P. Reichenbach, 1995. GIS Technology in Mapping Landslide Hazard, *Geographical Information Systems in Assessing Natural Hazards* (A. Carrera and F. Guzzetti, editors, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 135-175.
- Chung C.F., A.G. Fabbri and C.J. van Westen, 1995. Multivariate Regression Analysis for Landslide Hazard Zonation, *Geographical Information Systems in Assessing Natural Hazards* (A. Carrera and F. Guzzetti, editors, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 107-133.
- Chung C.F., and A.G. Fabbri, 1999. Probabilistic Prediction Model for Landslide Hazard Mapping. *Photogrammetric Engineering and Remote Sensing*. Vol. 65, No. 12 1999
- ESRI Inc, 1998. Environmental Systems Research Institute, Inc. Computer Software Version 7.2.1, Redlands, CA: ESRI Inc.
- Gallant J.C. and J.P. Wilson 1996. TAPES-G: A Grid-based Terrain Analysis Program for Environmental Sciences, *Computers & Geosciences*, Vol. 22, No. 7, pp. 713-722, 1996
- Gessler P.E., I.D. Moore, N.J. McKenzie, and P.J. Ryan 1995. Soil-landscape modelling and spatial prediction of soil attributes, *INT. J. Geographic Information Systems*, 1995, Vol. 9, No. 4, 421-432.

- Gorsevski P.V., P.E. Gessler, and R.B. Foltz 2000. Spatial Prediction of Landslide Hazard Using Logistic Regression and GIS, 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4):Problems, Prospects and Research Needs. Banff, Alberta, Canada, September 2 - 8, 2000.
- Johnson D.E., 1998. Applied Multivariate Methods for Data Analysis. Duxbury Press. An Intrnational Thomson Publishing Company 1998.
- McClelland D. E., R. B. Foltz, W. D. Wilson, T. W. Cundy, R. Heinemann, J. A. Saurbier, and R. L. Schuster 1997. Assessment of the 1995 & 1996 floods and landslides on the Clearwater National Forest, Part I: Landslide Assessment, *A Report to the regional Forester Northern Region U.S. Forest Service*, December 1997.
- Mark R.K., and S.D. Ellen 1995. Statistical and Simulation Models for Mapping Debris-Flow Hazard, Geographical Information Systems in Assessing Natural Hazards (A. Carrera and F. Guzzetti, editors, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 93-106.
- Montgomery D.R., and W.E. Dietrich 1994. A physically based model for the topographic control on shallow landsliding. Water Resources Research. Vol. 30. No. 4 Pages 1153-1171. April 1994.
- SAS Institute Inc., 1996 SAS/STAT Software: Changes and Enhancements Through Release 6.12. Cary, NC: SAS Institute Inc.
- 

## Authors

**Peter V. Gorsevski**, Ph.D. Candidate of Remote Sensing and GIS  
Department of Forest Resources, CNR, Moscow, Idaho. 83844-1133.  
Email: gors1571@uidaho.edu, Tel: +208-885-4946, Fax: +208-885-6226

**Paul Gessler**, Ph.D., Assistant Professor of Remote Sensing & GIS  
Department of Forest Resources, CNR, Moscow, Idaho. 83844-1133.  
Email: paulg@uidaho.edu, Tel: +208-885-2595, Fax: +208-885-6226

**Randy B. Foltz**, Ph.D., Research Engineer USDA Forest Service  
Rocky Mountain Research Station, 1221 S. Main St., Moscow, Idaho. 83843.  
Email: rfoltz@fs.fed.us, Tel: +208-882-3557, Fax: +208-883-2318